DOCUMENT RESUME

ED 213 748 TM 820 157

AUTHOR Forest, Laverne; Anderson-Smith, Charlotte
TITLE Are Evaluation Usage Types Related?
PUB DATE Oct 81
NOTE 14p.; Paper presented at the Joint Annual Meeting of the Evaluation Network and Evaluation Research Society (Austin, TX, September 30 - October 3, 1981).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Accountability; Communication (Thought Transfer); Decision Making; Learning; *Models; Motivation; *Program Evaluation; *Research Methodology
IDENTIFIERS *Evaluation Utilization

ABSTRACT

Research methods and findings relative to improving understanding of program evaluation usage patterns are reported. This study further refines an evaluation usage model consisting of five basic evaluation usage types: communication, learning, motivation, accountability and decision-making. This usage concept has been expanded to include importance and frequency as additional variables. Preliminary analysis supports the premise that usage types of less complexity, such as communication, learning and motivation, occur more frequently than those of higher complexity, such as accountability and decision-making. Conversely, decision-making in particular was determined to be of more importance than the other four types but utilized less often. An examination of patterns of evaluation usage was also done to determine if a hierarchy exists among the five usage categories. For example, do certain types of usage occur consequentially or simultaneously; can a usage taxonomy be developed and if so, is it monolithic or branching? Statistical analysis was used to examine if a logical hierarchy of usage is supported empirically. (Author/GK)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED213748

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Are Evaluation Usage Types Related?*

Introduction and Background

Program evaluations (used in the broadest sense to include curriculum analysis, project evaluation, program review, evaluation research, and other similar terms) are not new. Curriculum studies trace back to the 1930s. However, since 1964 the field has grown to the point that any state or federal legislation relative to education does not pass without a section dealing with expected formal evaluation. G.A.O.'s Annual Source Book of Federal Program Evaluation now lists evaluations totaling over $1,000,000,000/year. Whether conducted internally as part of a project or school curriculum or conducted by an external contractor, these formal evaluations invariably run $100,000 or more. For instance, University of Wisconsin-Extension is annually spending over $200,000 on comprehensive reviews when time of faculty involved is considered.

The proliferation of these formal evaluations and the dollars invested in them would indicate they are extremely useful and valuable. In fact, until about 1970, very few researchers and evaluators questioned the usefulness of program evaluations.

However, a paradox exists; evaluators say evaluations are underutilized. For instance, Fawcett and Wright (1980) say the "most frequently voiced concerns of evaluation researchers is that their work is underutilized and consequently has little impact." p. 1160. House (1973) says "how many days, fresh from his data, the evaluator presents his work, only to be received with transparent disappointment or mock enthusiasm." p. 256. Brown, Newman and

*Prepared by Laverne Forest and Charlotte Anderson-Smith and presented at 1981 Joint Meeting of Evaluation Network and Evaluation Research Society, Austin, Texas, October 1981.

TM 820 157

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Laverne Forest
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Rivers (1980) state "educators have become increasingly concerned with the utility of their evaluation activities and the effect that their conclusions and recommendations have on decision making." p. 67.

Sjogren (1973) says "the situation is almost a comedy. An evaluator is hired to play the role while both he and the project director know that the evaluation activity will make little real difference to anybody. While the evaluation role is being played, the real evaluations that do make a difference are being made by the project staff and the project constituents." p. 267.

Research Review on Utilization

Several empirical studies support these perceptions of evaluation users and usage. Caplan (1977) found program evaluations to be used as a knowledge source in 20% of the reported decision making instances by 204 interviews of "persons holding important positions in various departments, major agencies, and commissions of the executive branch of the U.S. government." p. 352.

Fox (1976) reports on the degree to which the U.S. Senate uses program evaluation as a technique for oversight. His summary indicates that program evaluation is sixth on the list as a source of information and data by Senate committee chairs and designated committee staff. Andringa (1976) in an informal survey of congressional staff found evaluation 11th in a list of 11 factors which influence legislation. Florio, Behrman, and Goltz (1980) found the Research and Development Community 10th out of 13 in providing the most useful information for determining legislative policy in education.

Nachmias (1980) summarizes the concerns saying "the evaluation community is undergoing a crises of unfilled expectations. There are only a few documented cases in which evaluation dramatically and immediately impacted directly on policy decisions." p. 1165.

The apparent underutilization of evaluation has led many to speculate why, and the reasons given vary widely. Bernstein and Freeman (1975) upon analyzing 382 evaluations said that one-half of them were deficient in design, sampling, or validity measures. p. 97. Nachmias offered that there was too much variability in quality of information. Sjogren (1973) said many evaluations have provided too little, too late. Rein and White (1980) said "social scientists and policy makers live in different worlds, characterized by conflicting values, rewards, and language." p. 119. Chelimsky (1980) said much evaluation research has low credibility. Banks and Clark (1980 argued evaluation must be done within the budgetary cycle so that they're more timely.

Florio, Behrmann, and Goltz (1980) provided a number of ways which U.S. Congressional staff felt could improve evaluations. They include better timing, more relevant information, conciseness, concrete recommendations, visuals and charts, and early direct involvement with users.

Rehka Agarwalla-Rogers (1977) concluded evaluations could be better utilized if program officials participated in the evaluation process and by appointing evaluators who are insiders to the organization being evaluated. Reasons she listed for nonusage of evaluation were:
(1) conflicting interests, (2) lack of participation by administrators, (3) lack of mutually agreed upon problem definition, (4) lack of effective liaison between administrators and evaluators, (5) lack of emphasis on problem solutions, (6) over emphasis on negative aspects, and (7) untimely feedback.

Berke (1980) found that the executive branch of government uses evaluation differently than the legislative branch. Usage in Congress was less dramatic and observable. The executive branch relied heavily on evidence from evaluation. Stewart (1979) essentially found the same thing in studying the appropriations process for the Cooperative Extension Service. Florio et al (1979) found that staffers had varying information needs.

Francis, King, and Riddlesperger (1980) found that administrators within agencies had different expectations of evaluation than did legislators. Results of many studies suggest that a basic reason for nonusage is possible irrelevancy of evaluation evidence due to gathering the wrong data.

Campeau (1980) concluded that certain strategies lead towards increased usage of evaluation: (1) arrange for early and continuing interaction with policy makers and implementors, (2) focus data analyses on gaps, (3) arrange for crucial, face-to-face briefing with official and time them prior to decisions, (4) bring in corroboration of findings if possible, (5) include in briefings both policy makers and implementors, (6) use a workshop format for briefings.

Mendelsohn and Bosner (1980), like Campeau and Agarwalla-Rogers, suggested a stronger relationship and bridging occur between managers of programs and evaluators at the working level to improve usage of evaluation by program managers.

However, Dickey's (1980) empirical study showed decision-maker attitude, evaluation timeliness, decision-makers' perceptions of procedures, and helpfulness as the only independent variables related to utilization. Surprisingly, she concluded the collaborative, active social involvement by the decision makers has not related to higher utilization.

Several researchers are beginning to uncover another reason for the apparent underusage, however. They have realized that evaluation does not often have an immediate impact on policy decisions and also that decisions are seldom made at one dramatic point in time but instead evolve incrementally.

Patton (1978), one of the early leaders in these discoveries, found "the more typical impact was one where evaluation findings provided additional pieces of information in the difficult puzzle of program action, permitting some reduction in the uncertainty within which any federal decision-maker

inevitably operates." p. 28. In his study, Patton found 78% of decision makers in 20 different evaluations feeling evaluation had had an impact, and that 90% of evaluators involved in those 20 projects felt that evaluation had impact. None of the impacts, however, were described as being direct and immediate leading to major concrete program decisions.

Perrin (1980) in analyzing these types of findings said:

> studies exploring the question of research utilization have indicated that research and evaluation generally are not used directly in an instrumental sense for the making of specific decisions. When it is used directly, it is for minor program adjustments rather than for larger policy changes. This has been taken by many to indicate the lack of use of evaluation and research in policy making. However, more recent evidence, as well as reanalysis of other data, suggest that evaluation is used but rather in a conceptual or "enlightenment" sense. . . .Evaluation use is an indirect and defuse process, providing orientation for issues rather than solutions for problems. Utilization is more complex and amorphous than originally conceived. . . .it is difficult or impossible in most cases to point a finger at any one person and identify that individual as the one responsible for any given decision. . . .there is an interesting paradox here. Most work and analysis on an issue is done by individuals lower down the chain of command. Higher-level managers are generally involved in approval or rejection of ideas. p. 1-5.

Weiss (1977) says:

> government officials use research less to arrive at solutions than to orient themselves to problems. They use research to help them think about issues and define the problematics of a situation, to gain new ideas and new perspectives. They use research to help formulate problems and to set the stage for future policy actions. p. 534.

It appears, then, that evaluation is not only used for immediate decision making but also for enlightenment or learning and developing a better understanding of the program.

The Research Problem

The above studies and analyses have contributed significant findings in the study of program evaluation utilization. These studies have opened the question of whether the boundaries or essence of evaluation utilization are really known. Utilization is apparently a much broader and more complex concept than just mere direct, immediate, tangible input to program related decision-making.

Presuming a broader concept of program evaluation usage, the studies by Patton (1978), Alkin et al (1974), and Roberts and Larsen (1972) tend to show a more extensive usage of evaluations than previously thought. Forest and Marshall (1981) in fact found five types of general usages to exist and to be easily recognized by evaluators and users of evaluation. These usage types were: sharing and increased communication, increased understanding, increased motivation, input to decision making, and input to accountability and reporting.

Perhaps then a key reason for underusage is the lack of an adequate concept or definition of evaluation usage. Palumbo and Wright (1980) concluded "whether evaluation research is utilized depends on what is meant by the term utilized." Smith (1980) specifically said "most writers do not define what they mean by evaluation use. . . ." p. 24. Patton (1978) said "the predominate image of nonutilization that characterizes much of the commentary on evaluation research can be attributed in substantial degree to a definition of utilization that is too narrow in its emphasis on seeing immediate, direct, and concrete impact on program decisions." p. 32.

The research problem studied and reported on here then is this latter point; evaluators do not yet know what "evaluation usage" actually means. The authors are convinced that usage does entail the notions of being practical, being pragmatic, making a difference, and leading to consequences, but any number of practical consequences have been shown to be possible. More specifically, this paper reports research on the degree to which these various consequences or usages are empirically related.

Research Hypotheses

The five types of evaluation usage by Forest and Marshall (1981) were further studied and analyzed. The Forest/Marshall paper suggests the usage

categories are inversely related as to frequency and importance; that is, learning, motivation and communications occur more frequently than accountability and decision-making but are rated as being less important. This observation suggested a hierarchy of evaluation usage exists, similar to the hierarchies of learning proposed by Bloom (1956), Krathwohl et al (1964), and Gagne (1970).

The research hypothesis tested statistically was:

There is no hierarchical relationship between the five usage types: communication (C), increased learning and understanding (L), increased motivation (M), decision-making (D), and accountability (A).

Methodology

This study used data from the Forest and Marshall (1981) study. These data were developed as follows. Approximately 300 general statements, each representing a potential usage type, were examined by an "expert" panel of sixteen people, all with experience in evaluation theory and practice. Among the tasks performed by this panel was the categorizing of each statement into one usage type. The criterion established for accepting a statement as representing a particular type of usage was 75% agreement among panel members. Fifty-four statements met this criterion and were used in a pilot questionnaire. A final 27 statements were chosen for a questionnaire which was mailed to recipients of the Shawano Impact Evaluation (Forest and Marshall, 1981), a study done in Shawano County, Wisconsin. Of a population of about 1,300 recipients, a stratified sample of 265 received the questionnaire. One hundred eighty people (68%) returned the survey. Their responses to 25 of the statements, scored dichotomously as either "yes" or "no", form the data base for this study.

Analysis

First, Pearson's correlation between the usage types found decision-making correlated with all four of the other types, while accountability was related to all but motivation. Motivation was only correlated with decision-making, while learning and communication correlated positively with each other, accountability and decision-making. Two possible hierarchies which might explain these correlations would be the following:

Case 1 Case 2

Figure 1

Based on the frequency of positive responses, the rating of importance given by the panel, and deductive reasoning, two other possible hierarchies were identified:

Case 1 Case 2

Figure 2

Both hierarchies found a branching occurring before accountability/ decision-making. A great deal of uncertainty exists, however, concerning the relative placements of learning, motivation and communication.

A procedure developed by Baker and Hubert (1977) tests whether or not a particular hierarchy exists among a set of dichotomously scored items. The Quadratic Assignment Program (Baker, Hubert, and Schultz, 1977) makes the necessary calculations.

Using the Quadratic Assignment Program

Six distinct steps are followed in applying the Quadratic Assignment Program to a particular data set.

1. Designing the theoretical or structure matrtix. A matrix of data items is set up in a sequence of logical progression; that is, a positive response to item 1 logically precedes a positive response to item 2. A one is assigned to a particular cell within the matrix if the row item logically precedes the column item; otherwise a zero is assigned to that cell. All items on or below the diagonal are assigned zeros.
2. Developing the initial data or raw proximity matrix. The first computer run will count the number of people who scored items in a "disconfirmatory" manner; that is, they scored a positive response on a column item and a negative response on the corresponding row item. The raw scores printed in the upper triangle above the diagonal are considered in Step 3. The scores printed below the diagonal are of no further interest.
3. Applying the decision rule. A decision is made by the researcher regarding the maximum number of disconfirmatory responses which are acceptable. This value is called "epsilon" and is made by choosing

several possible values based on the particular data set and subjecting them to Step 4. The epsilon which provides the best "fit" of the proximity matrix to the structure matrix is selected.

4. Developing the proximity matrix. Once epsilon has been selected, a proximity matrix is developed by assigning a one to all cells above the diagonal in which the raw data score is less than or equal to epsilon. All other cells, including those on and below the diagonal, are assigned a zero. This matrix is analogous to the structure or theoretical matrix.

5. Calculating the statistic "gamma". The computer program cross-multiplies the values in the cells of both the structure and proximity matrices. If a one is present in both, the product will, of course, be one. A zero in either or both matrices will result in a product of zero. These products are then summed to calculate the statistic "gamma".

6. Determining statistical significance. To determine whether or not the gamma for a particular data set could have occurred by chance, the QAP then randomly permutes the rows and columns of the proximity, or data, matrix through all possible combinations, up to 1000 times in the case of large matrices. The statistic gamma is calculated for each permutation forming a sampling distribution of gamma. [This is analagous to Fischer's Exact Text, Fischer (1935).] A significance level can then be chosen, and the gamma obtained from the original data matrix can be compared to the distribution. If the value lies high enough in the distribution, at the 0.01 level for example, the evidence would be quite strong that the data did not occur in a random fashion but resembled strongly the structure of the theoretical matrix.

In this study, a theoretical matrix was set up, 25 by 25, representing the original items used in the Forest and Marshall paper.

Results

The data were analyzed for two of the four theoretical hierarchies (see Figure 2, Case 1 and Case 2). The gamma value for Case 1 was 188 and the value for Case 2 was 169. These statistics were both significant at p = .001. The null hypothesis was therefore rejected.

Conclusions and Implications

Due to the relatively large number of statements which measured each usage category, all inexactness was impossible to eliminate. The results did indicate, however, that evaluation usage does not occur at random and that some ordering of usage events exists. Future research should continue to test this order of usage.

An understanding of evaluation usage types, their order of occurrence, and the relationship of that order to time sequences will allow program evaluations to be designed and implemented which take advantage of the naturally occurring sequence of usage events. More accurate and reasonable criteria for "meta-evaluation", or the critique of evaluation itself, can also be developed once the evaluative usage process is understood.

References

Agarwalla-Rogers, Rehka. "Why Is Evaluation Research Not Utilized?" in Evaluation Studies Review Annual, Marcia Guttentag (Ed.), Volume 2, 1977 pp. 327-333.

Alkin, M. C., Kosecoff, J., Fitz-Gibbon, C. and Seligman, R. "Evaluation and Decision-making: The Title VII Experience", CSE Monograph Series in Evaluation Center for the Study of Evaluation, University of California, Los Angeles, 1974.

Andringa, R. C. "Eleven Factors Influencing Federal Education Legislation" in S. Halperin and G. Kaplan (Eds.) Federalism at the Crossroads, Washington, D.C., Institute for Educational Leadership, 1976.

Baker, Frank B. and Hubert, Laurence J. "Inference Procedures for Ordering Theory", Journal of Educational Statistics, Autumn, 1977, Volume 2, No. 3, pp. 217-233.

Baker, Frank B., Hubert, Laurence J., and Schultz, James V. Quadratic Assignment Program, Laboratory of Experimental Design, 1075 Educational Science Building, University of Wisconsin-Madison, Madison, Wisconsin, March, 1977.

Banks, Jerry and Clark, Robert F. "Evaluation Research and the Budget Cycle", Policy Studies Journal, Volume 8, No. 7, Special No. 3, 1980, pp. 1195-1203.

Bernstein, I. L. and Freeman, H. E. Academic and Entrepreneurial Research, New York: Russell Sage Foundation, 1975.

Berke, Iris. Executive Branch vs. Congressional Use of Evaluation, Presented at Evaluation Research Society Annual Meeting, Washington, D.C., November, 1980.

Bloom, B. S. (Ed.), Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook I: Cognitive Domain, New York: Longmans, Green, 1956.

Brown, Robert, Newman, Dianna and Rivers, Linda. "Perceived Need for Evaluation and Data Usage as Influences on an Evaluation's Impact", Educational Evaluation and Policy Analysis, September-October, Volume 2, No. 5, 1980, pp. 67-73.

Campeau, Peggie. Increasing the Usefulness of Federal Program Evaluation, Presentation at Evaluation Research Society Annual Meeting, 1980.

Caplan, Nathan. "Social Research and National Policy: What Gets Used; By Whom; For What Purposes; and With What Effects", Marcia Guttentag (Ed.), Evaluation Studies Review Annual, Volume 2, 1977, pp. 351-359.

Chelimsky, Eleanor. "Evaluation Research Credibility and the Congress", Policy Studies Journal, Volume 8, No. 7, Special No. 3, 1980, pp. 1177-1184.

Dickey, Barbara. "Utilization of Evaluations of Small Scale Innovative Educational Projects", Education Evaluation and Policy Analysis, November-December, 1980, Volume 2, No. 6, pp. 65-77.

Fawcett, Steve and Wright, Paula. "Evaluation", Policy Studies Journal, Volume 8, No. 7, Special No. 3, 1980, pp. 1160-1163.

Fischer, R. A. The Design of Experiments, Edinburgh, Oliver and Boyd, 1935.

Florio, David, Behrmann, Michael and Goltz, Diane. "What Do Policy Makers Think of Educational Research and Evaluation? or Do They?", Educational Evaluation and Policy Analysis, November-December, 1979, Volume 1, No. 6, pp. 61-81.

Forest, Laverne and Marshall, Mary. Impact of Extension in Shawano County: 3. Usage and Appraisal, Program and Staff Development, UW-Extension, Madison, Wisconsin, 1981.

Fox, Harrison. Legislative Evaluation, Proceedings of Second Annual Conference of Evaluation Network, St Louis, Missouri, 1976, pp. 15-26.

Francis, Wayne, King, James and Riddlesperger, James Jr. "Problems in the Communication of Evaluation Research to Policy Makers", Policy Studies Journal, Volume 8, No. 7, Special No. 3, 1980, pp. 1184-1194.

Gagne, R. M. The Conditions of Learning, (2nd edition), New York: Holt, Rinehart, and Winston, 1970.

House, E. "The Relevance of Evaluation" in Studies in Transactional Evaluation, Robert Rippey (Ed.), Berkeley, California, McCutcheon, 1973, pp. 256-266.

Krathwohl, D. R., Bloom, B. S., and Masia, B. B. Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook II: Affective Domain, New York: David McKay, 1964.

Mendelsohn, Robert and Bosner, Leo. The Liaison Role in Evaluation: Serving as as Bridge Between Managers and Evaluators, Presentation at Evaluation Research Society Annual Meeting, 1980.

Nachmias, David. "The Role of Evaluation in Public Policy", Policy Studies Journal, Volume 8, No. 7, Special No. 3, 1980, pp. 1163-1169.

Palumbo, Dennis and Wright, Paula. "Decision-making and Evaluation Research", Policy Studies Journal, Volume 8, No. 7, Special No. 3, 1980, pp. 1170-1177.

Patton, Michael. Utilization-Focused Evaluation, Beverly Hills, California: Sage, 1978.

Perrin, Burt. Evaluation as Enlightenment: Implications for the Utilization of Evaluation in Policy Development, Presentation at Evaluation Research Society Annual Meeting, 1980.

Rein, M. and White, S. H. "Can Policy Research Help Policy?", The Public Interest, 1977, Volume 49, No. 4, pp. 119-136.

Roberts, A. and Larsen, J. "Effective Use of Mental Health Research Information", Final Report, NIMH Grant No. R01-MH-15445, 1972.

Sjogren, Douglas. "Program Evaluation in Schools" in Studies in Transactional Evaluation, Robert Rippey (Ed.), Berkeley, California, McCutcheon Publishing Corporation, 1973, p. 267-274.

Smith, Nick. "Evaluation Utilization: Some Needed Distinctions", Evaluation News, Number 16, Summer 1980, pp. 24-25.

Stewart, David. "Cooperative Extension, the Carter Administration, and the 95th Congress: An Empirical Model of the Politics of Appropriation", Unpublished Dissertation, University of Wisconsin, Madison, 1979.

Weiss, Carol. "Research for Policy's Sake: The Enlightenment Function of Social Research", Policy Analysis, 1977, Volume 3, No. 4, pp. 531-545.